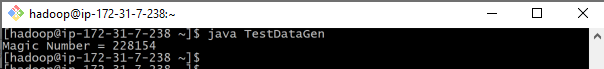
Rakshith Churchagundi Amarnath

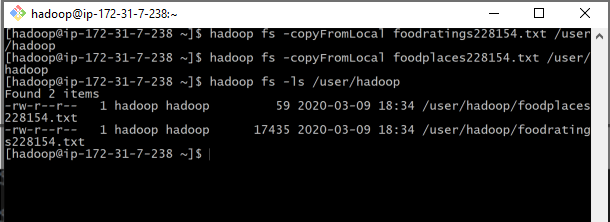
A20424771

CSP 554 – Assignment 8

Exercise 1:

Magic Number – 228154





Step C:

>>> from pyspark.sql.types import \*

>>> ratings=StructType(

... [

... StructField("name",StringType(),True),

... StructField("food1",IntegerType(),True),

... StructField("food2",IntegerType(),True),

... StructField("food3",IntegerType(),True),

... StructField("food4",IntegerType(),True),

... StructField("placeid",IntegerType(),True),

... ])

>>>foodratings=spark.read.schema(ratings).csv('hdfs:///user/hadoop/foodratings228154.txt')

>>>foodratings.printSchema()

>>>foodratings.show(5)



Exercise 2:

>>> from pyspark.sql.types import \*

>>> places=StructType(

... [

... StructField("placeid", IntegerType(), True),

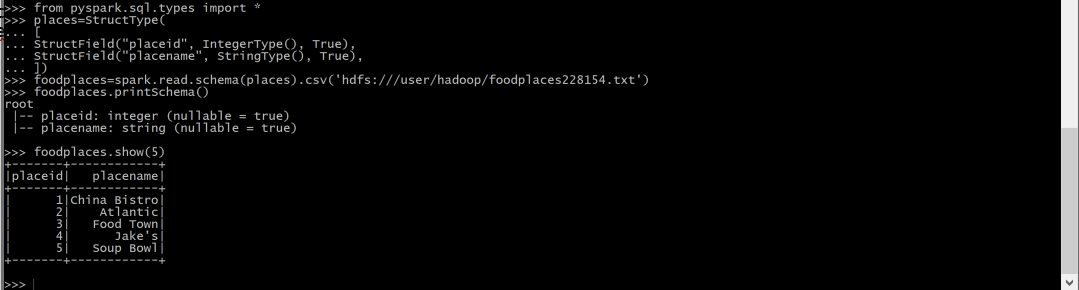
... StructField("placename", StringType(), True),

... ])

>>> foodplaces=spark.read.schema(places).csv('hdfs:///user/hadoop/foodplaces228154.txt')

>>> foodplaces.printSchema()

>>> foodplaces.show(5)



Exercise 3:

Step A:

>>> from pyspark.sql.types import \*

>>> foodratings.createOrReplaceTempView("foodratingsT")

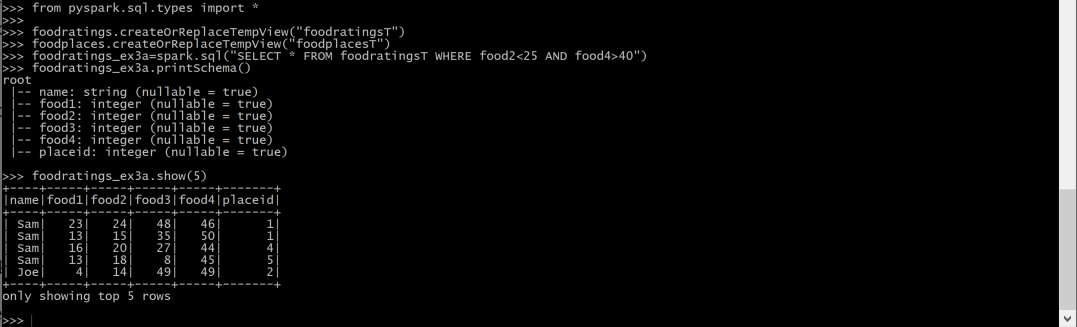
>>> foodplaces.createOrReplaceTempView("foodplacesT")

Step B:

>>> foodratings\_ex3a=spark.sql("SELECT \* FROM foodratingsT WHERE food2<25 AND food4>40")

>>> foodratings\_ex3a.printSchema()

>>> foodratings\_ex3a.show(5)



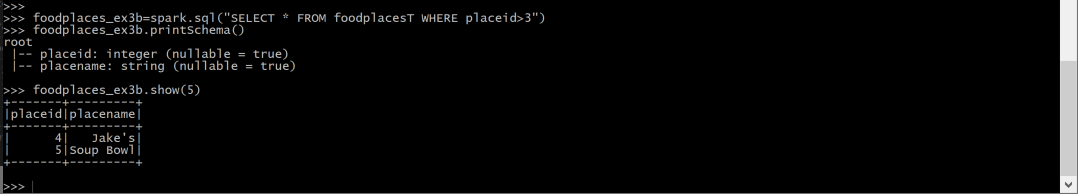
Step C:

>>>

>>> foodplaces\_ex3b=spark.sql("SELECT \* FROM foodplacesT WHERE placeid>3")

>>> foodplaces\_ex3b.printSchema()

>>> foodplaces\_ex3b.show(5)

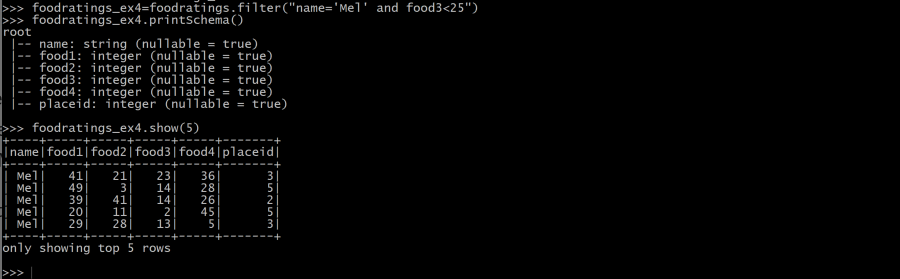


**Exercise 4:**

>>> foodratings\_ex4=foodratings.filter("name='Mel' and food3<25")

>>> foodratings\_ex4.printSchema()

>>> foodratings\_ex4.show(5)

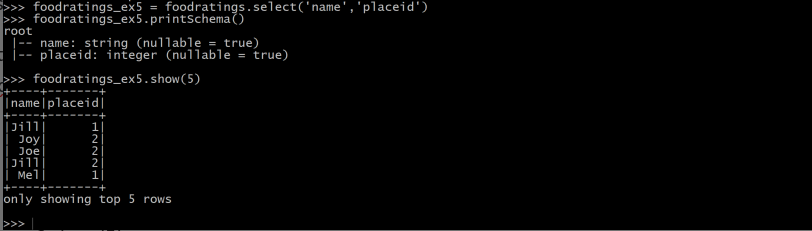


**Exercise 5:**

>>> foodratings\_ex5 = foodratings.select('name','placeid')

>>> foodratings\_ex5.printSchema()

>>> foodratings\_ex5.show(5)



**Exercise 6:**

>>> ex6 = foodratings.join(foodplaces,foodratings.placeid==foodplaces.placeid)

>>> ex6.printSchema()

>>> ex6.show(5)

